Case Study

Causal analysis to determine the strength of effect of Cephalosporin use on antimicrobial resistant infections

incubate.bio

Dr Raminderpal Singh
raminderpal@incubate.bio

In collaboration with

Department of Process Engineering
Stellenbosch University

Abstract

Antimicrobial resistant (AMR) infections are a global threat to public health. The use of the antibiotic, Cephalosporin, has been associated with the emergence of resistant organisms. Furthermore, the rise in resistance against Cephalosporin is of great concern to the WHO. The aim was to determine the strength of the causal effect of Cephalosporin use on the incidence of resistant infections. We used a cohort of data consisting of antibiotic use in human healthcare and agriculture, the levels of antibiotics and resistant gene fragments in wastewater, and AMR incidence reports for this case study.
The causal effect of Cephalosporin is quantified for 15 pathogenic microbes on the WHO's priority watchlist. We find that Cephalosporin use in agriculture has the strongest causal effect on the resistant infections of eight of the microbes. The Cephalosporin use in human healthcare has the strongest causal effect on the resistant infections of the remaining microbes, save for one, of which the strongest causal effect is re-infection. This analysis shows evidence of a strong causal association between the antibiotic use in agriculture (veterinary treatment, prophylactic use, etc.) and antimicrobial resistant infections.
An explanation of the method used here can be found in the Addendum.

Background for the case study

Antimicrobial resistance is a growing concern affecting both high-income and low- to middle-income countries. AMR is not only a burden on public health, but also affects the agricultural sector considering the zoonotic nature of many of the microbes involved. Therefore AMR is studied in the context of One Health (Figure 1) - human health, animal health and environmental health [1]. The interactions between these ecosystems can be translated to feedback loops and reinforcements in terms of causal effects.

There are a number of factors that can contribute to the spread of AMR:

- Rapid natural microbial evolution
- Horizontal gene transfer between microbes
- Excessive and inappropriate use of antimicrobials in healthcare
- Prophylactic use of antimicrobials in agriculture
- The presence of AMR gene fragments in bodies of water

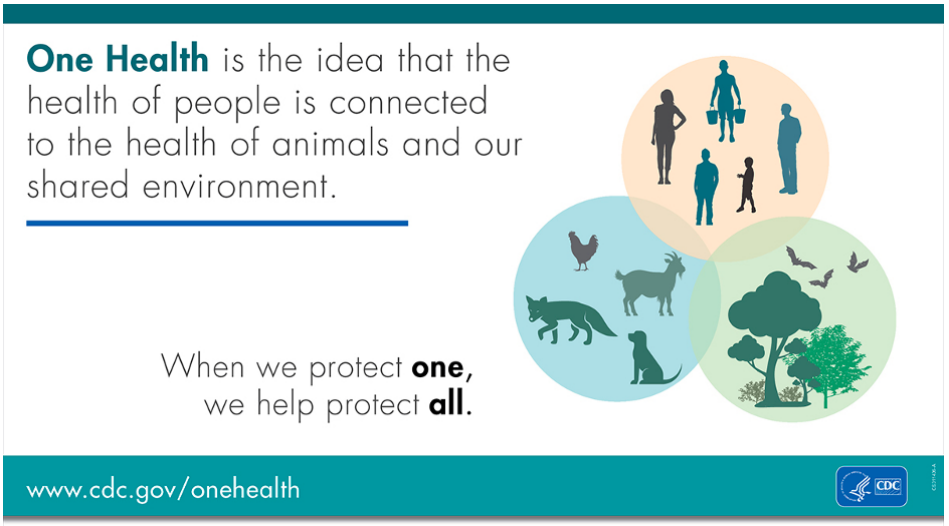

Figure 1: One Health - the health of humans, animals, and the environment is shared [2].

Understanding which of these factors are main drivers of AMR incidence and the causal effect behind these mechanisms will be key in targeting the spread of AMR. Current efforts however are largely focussed on correlative and machine learning analyses. These types of analyses identify trends and are useful for AMR surveillance and prediction, but give little insight into the main causal mechanism behind AMR incidence. Evidence as to which factors are the main drivers of resistant infections is needed before any intervention, such as changes in policy, can be executed.

One antimicrobial, against which resistance is of significant concern, is Cephalosporin. Cephalosporin resistance has shown to be associated with a variety of hospital acquired microbial infections [3], often referred to as “superbugs”. The occurrence of this type of resistant infections has been on the increase in the last few decades [3]. The incidence of Cephalosporin resistant infections is of great concern to the WHO [1].

This case study shows how we quantified the causal effects of Cephalosporin on AMR infections in Stellenbosch, South Africa - insights that were not clear from correlative and machine learning analyses alone.

Need for causal analyses

The machine learning analyses from the study by Louw *et al.* [4] show that the antibiotic levels in wastewater are the best predictors for surveillance and prediction of resistant infections. Although these antibiotic features result in accurate machine learning models for the prediction of AMR incidence, there is no overlap between these features and the features that show strong correlative links with AMR incidence, nor with the features that have strong causal effects on AMR incidence.

This shows that correlative and machine learning analyses alone are lacking for finding the main causal effects of AMR incidence. Causal analyses, such as the work described by Pearl [5, 6, 7], bridges this gap. Without causal analyses, only a limited subset of the questions surrounding resistant infections and One Health can be answered.

Correlation analyses (Figure 2) alone cannot reveal any of the underlying causal structure in the data. Machine learning analyses identified the antibiotic levels in wastewater samples as the best predictors of resistant infections [4].

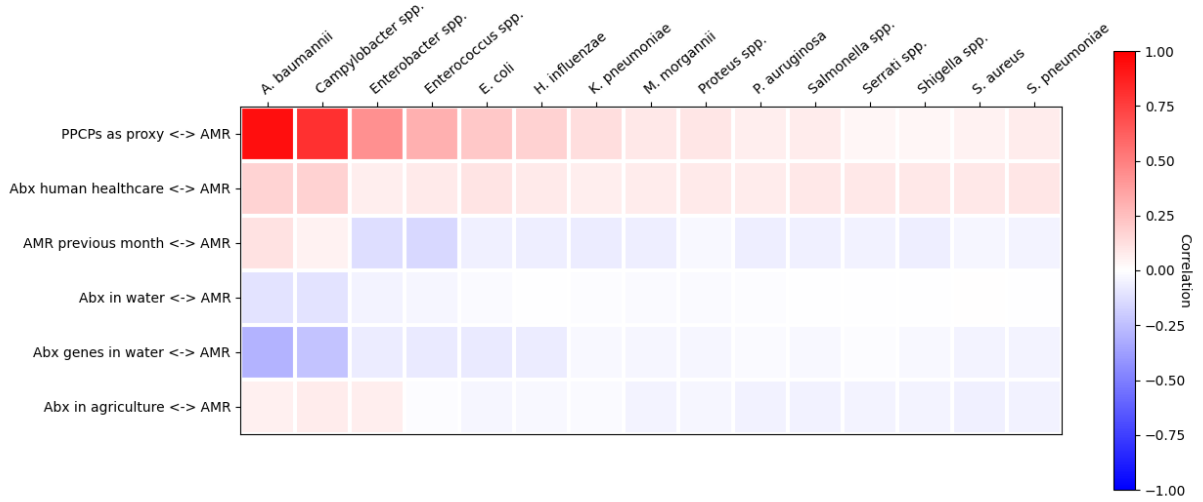

Figure 2: Correlation between potential drivers of AMR and the incidence of AMR (y-axis) across all microbes (y-axis). Positive correlations are shown in red and negative correlations are shown in blue

Causal question

Antimicrobial resistance is a complex question and we therefore assume that the main drivers of AMR infections might be different for the various types of resistant infections. We consider all of the microbes on the WHO's watchlist [8]. We hypothesise that there are six main potential drivers for resistant infections. These drivers are selected based on the findings and assumptions of studies by Louw *et al.* [4] and Archer *et al*. [9], the main data sources of this case study.

These drivers are represented as nodes in Figure 3, which is a directed acyclic graph (DAG) and is used for hypothesis testing in this case study. The potential drivers with the corresponding nodes on the DAG (Figure 3) are:

- Changes in seasonality and population size (pharmaceuticals and personal care products (PPCPs) as proxy)*
- The use of antibiotics (Cephalosporin) in human healthcare (Abx human health)
- The use of antibiotics (Cephalosporin) in agriculture (Abx agricultural use)
- The presence of antibiotics (Cephalosporin) in wastewater (Abx in water)
- Antibiotic resistant gene fragments in wastewater (AMR genes in water)
- The current AMR incidence levels in the community driving re-infection (AMR in humans previous month)

*The levels of pharmaceutical and personal care products detected in wastewater samples are used as a proxy for the effect of the fluctuation in population size, which in itself is affected by seasonality. The population of the town fluctuates throughout the year as the large population of students and seasonal workers return home during the holidays.

Causal question addressed in this case study: considering all the potential causes of resistant infections in the context of One Health, what is the strength of the causal effect of Cephalosporin on AMR infections?

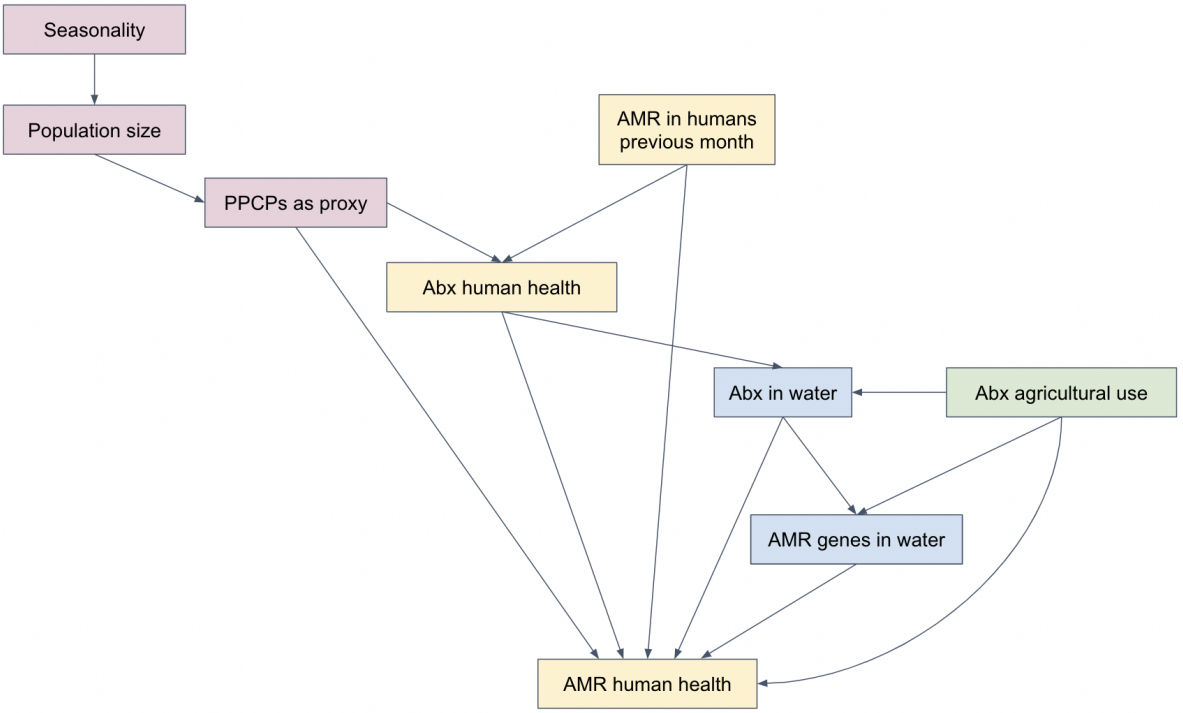

Figure 3: DAG for the causal inference of AMR incidence in humans.

Results of causal analysis

The DAG in Figure 3 was used as the hypothesised causal structure for this case study. This DAG structure remained the same across all scenarios (DAGs) tested for the individual microbes.

Causal effects are shown in Figure 4:

1. **Across all microbes, the use of antibiotics in agriculture either has a strong negative or strong positive causal effect. See Figures 4 and 5.**
2. AMR gene fragments in wastewater, which is sometimes hypothesised to be a causal effector of resistant infections, show very low causal effects on AMR infections.
3. A similar low causal effect of fluctuation in population size (PPCPs) on resistant infections is seen across all microbes.
4. The use of antibiotics in human healthcare is also very low across all microbes, save for *Campylobacter* spp., *E. coli*, and *Salmonella* spp., on which this feature has a strong causal effect.
5. The antibiotic levels in wastewater have very low to strong negative causal effect on resistant infections across all microbes.
6. A similar effect is seen across all microbes for the causal effect of the previous month's infections on the resistant infections of the current month, save for *Enterococcus* spp. where this feature has a strong causal effect on resistant infections.

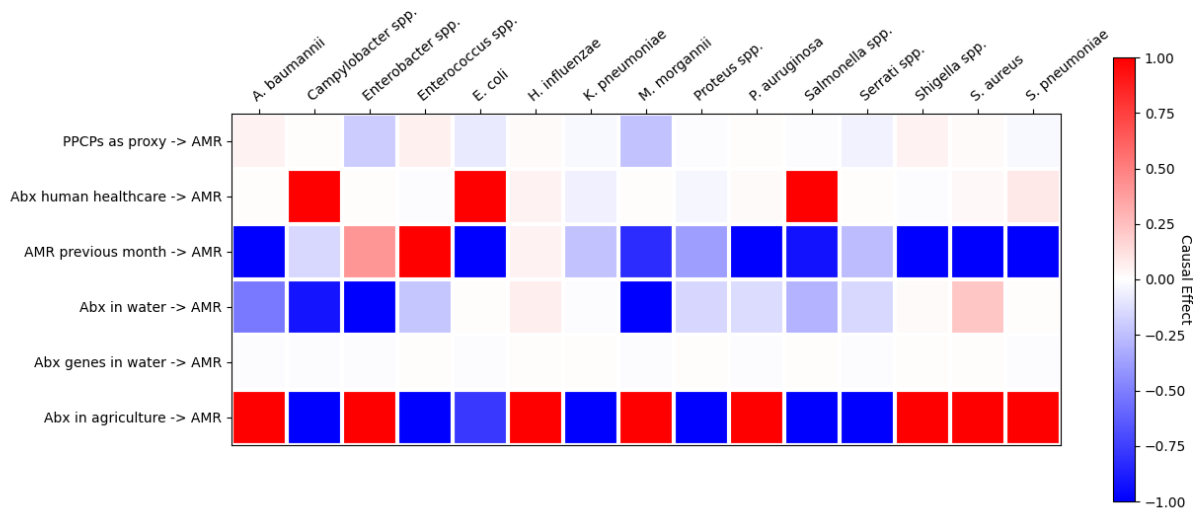

Figure 4: Causal strengths between entities (y-axis) of the DAG in Figure 3 across all microbes (y-axis). Positive causal effects are shown in red and negative causal effects are shown in blue.

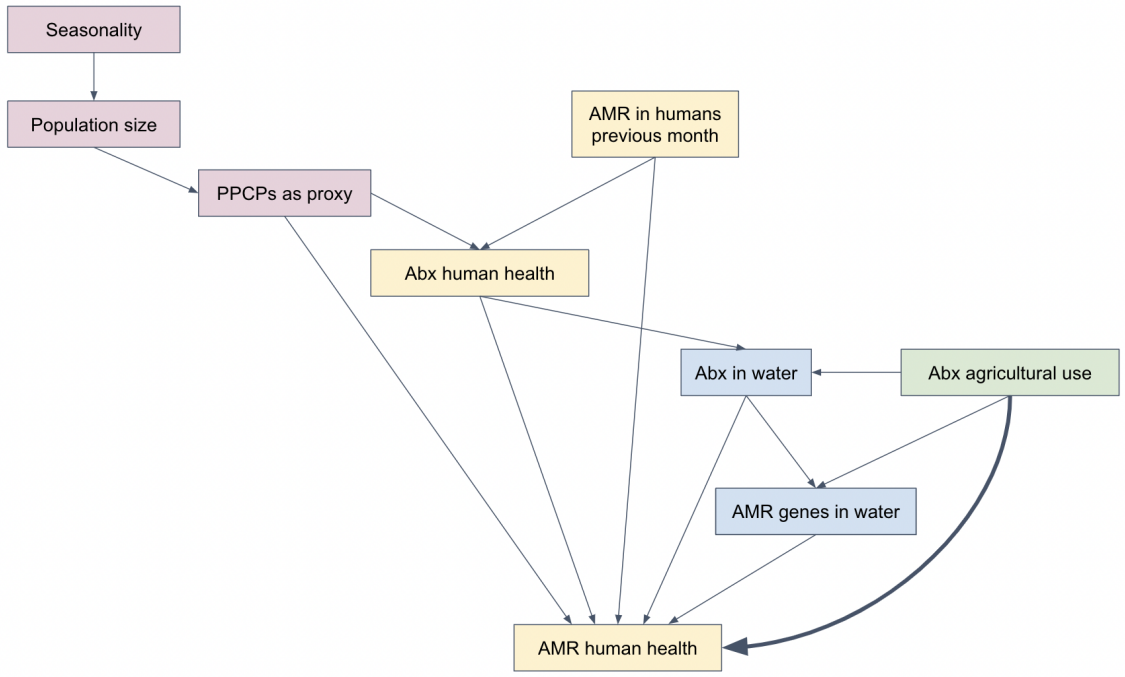

Figure 5: DAG with the largest causal effect across all microbes shown by the bold arrow.

Overall, causal analysis (Figures 4 and 5) shows that the agricultural use of Cephalosporin has the strongest causal effect on the resistant infections for the following eight microbes:

- *A. baumannii*
- *Enterobacter* spp.
- *H. influenzae*
- *M. morgannii*
- *P. aeruginosa*
- *Shigella* spp.
- *S. aureus*
- *S. pneumoniae*

In addition, we see that use of Cephalosporin in human healthcare has the strongest causal effect on the resistant infections of *Campylobacter* spp., *E. coli*, and *Salmonella* spp., while the resistant *Enterococcus* spp. infections of the previous month have the strongest causal effect on the current resistant *Enterococcus* spp. infections.

Impact of these results

To reduce the overall incidence of AMR infections in Stellenbosch, the most effective intervention to target would be the restriction of antibiotics in agriculture.

Data

Data collected by the ReNEW [9] and Grand-Challenges Africa - Antimicrobial resistance [4] projects were used in this case study.

Acknowledgement

This work would not have been possible without the investigators of the ReNEW project [9]. We would also like to acknowledge the investigators of the Grand Challenges Africa - Antimicrobial Resistance project for collaborating on this case study [4].

Disclaimer

The views expressed in this document are those of Incubate.bio and do not necessarily reflect the views of the Grand Challenges Africa team or the ReNEW team, who collected the UWP data.

References

[1] World Health Organization, (2021). Global antimicrobial resistance and use surveillance system (GLASS) report: 2021 https://www.who.int/publications/i/item/9789240027336.

[2] Centers for Disease Control and Prevention, “One Health”, https://www.cdc.gov/onehealth/images/social-media/what-is-one-health-fb.jpg

[3] S. J. Dancer, (2001). The problem with cephalosporins, Journal of Antimicrobial Chemotherapy, Volume 48, Issue 4, Pages 463–478.

[4] C. Louw, E. Bester, E. Archer, F. Kamper, G. M. Wolfaardt, T. M. Louw, (2022), Wastewater-based epidemiology as a viable minimum data strategy for antimicrobial resistance surveillance (manuscript in progress).

[5] Pearl, J., & Mackenzie, D. (2018). The book of why: the new science of cause and effect. Basic books.

[6] Glymour, M., Pearl, J., & Jewell, N. P. (2016). Causal inference in statistics: A primer. John Wiley & Sons.

[7] Pearl, J. (2009). Causality. Cambridge university press.

[8] World Health Organisation. Media Centre, “WHO publishes list of bacteria for which new antibiotics are urgently needed.”, http://www.who.int/mediacentre/news/releases/2017/bacteria-antibiotics-needed/en/

[9] Archer, E., Holton, E., Fidal, J., Kasprzyk-Hordern, B., Carstens, A., Brocker, L., et al. (2022) Addressing Environmental- and Public Health through Urban Water Profiling of Emerging Contaminants in a South African Urban Setting. ChemRxiv: https://doi.org/10.26434/chemrxiv-2022-5gk0z.

Addendum: Causal analysis overview

Introduction to causal analysis

Based on Bayesian methods and mathematical logic, causal analysis aims to discover new evidence behind a causal system [1, 2, 3].

- How is it implemented? Through a network diagram showing which entities in a system affect others, based on current knowledge.
- What does it do? Using established statistical methods, the diagram is analysed to gain an understanding of cause-and-effect relationships appearing in measurements of and interventions in the system.
- Why do we need it? It leads to i) the uncovering of and evidence behind causes that explain the observed behaviour, and ii) identification of parts of the system that can be modified to change outcomes.

Statistical analysis (correlation)	Correlation + Causation
Difficult (to impossible) to determine which is the cause and which is the effect	It is clear which action is the cause and which is the effect, through **directionality** in and **quantification** of the relationships
Not possible to identify hidden variables/ confounders in a system	The presence of **hidden variables/ confounders** can be discovered and quantified
Difficult to catch spurious correlations	Easy to catch **spurious correlations** which may have no causality
Cascading effects across nodes and between systems cannot be understood	Quantification of **upstream and cross-stream** effects possible

Implementation in this case study

To answer the main hypothesis, which of the potential drivers of AMR incidence in humans are main risks (causes) of infection, the effects of the six potential causes of AMR incidence were quantified. From Figure 3, these direct causal effects are: *PPCPs as proxy*, *Abx human healthcare*, *Abx in agriculture*, *Abx in water*, *AMR genes in water*, and *AMR previous month*. The other causal effects in the DAG were also quantified, such as the effect of *Abx human healthcare* on *Abx in water*.

The causal effects were quantified by marginalising over confounding effects with the use of the backdoor criterion [2, 3], where the effect of one variable on another variable (the outcome) is quantified by marginalising over all other variables creating paths that are not direct effects. The directed acyclic graph (Figure 3) can be represented as a graphical model (Figure A1) and a set of structural model equations (Equations 1 - 7). Here the variables S, T, V, W, X, Y, and Z represent the nodes *PPCPs as proxy*, *Abx human healthcare*, *AMR previous month, Abx in water, Abx in agriculture*, *AMR genes in water*, and *AMR in humans* respectively.

The structural model shows the direct causal effects of the variables on one another, as well as the effect of unmeasured exogenous variables (U_i). These direct causal effects are also referred to as the average causal effects (ACE) and are represented in the graphical model (Figure A1) as the variables a, b, c, d, e, f, g, h, i, j, k, and l on the graph edges. Linear regression is used to quantify the direct causal effects (Equations 8 - 19) between the variables and the outcomes, thus the regression coefficients are the pathway coefficients (direct causal effects) [2]. The total causal effects can be written as Equations 20 - 23 with substitution. The indirect causal effects (Equations 24 - 27) can be calculated by subtracting the direct effects from the total causal effects. Computational analyses were run using Python 3.8 and the DoWhy package for causal inference [4, 5].

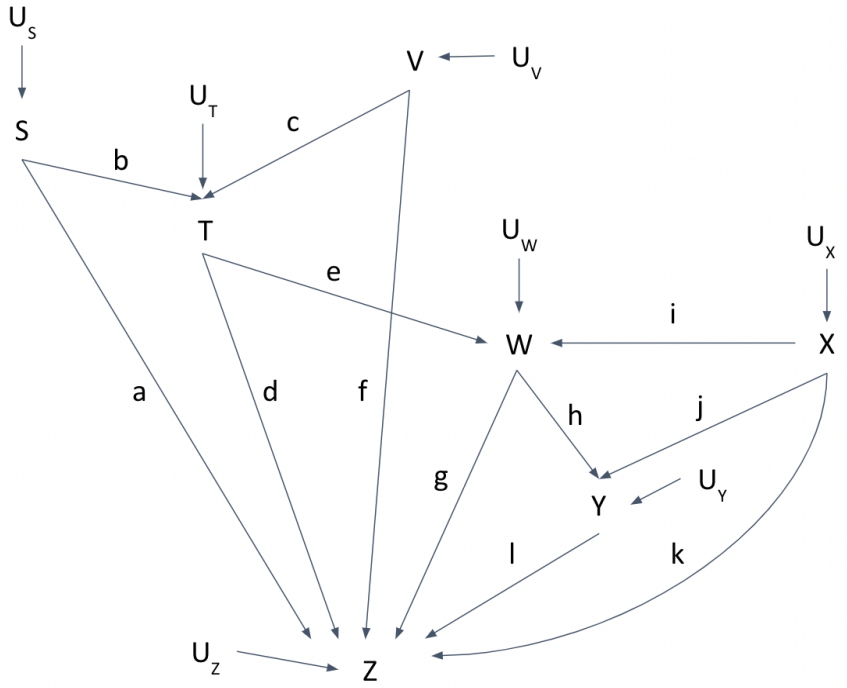

Figure A1: Graphical model of the directed acyclic graph (Figure 3).

Structural model

$S = U_S$
$T = bS + cV + U_T$
$V = U_V$
$W = eT + iX + U_W$
$X = U_X$
$Y = hW + jX + U_Y$
$Z = aS + dT + fV + gW + lY + kX + U_Z$ Equations 1 - 7

Direct causal effects

The effect of S on T:

$ACE_{S \text{ on } T} = E[T|S = s_n + 1, V = v_n] - E[T|S = s_n, V = v_n]$ where n = time
$= r_0 + b(S + 1) + cV - (r_0 + bS + cV)$
$= b$

Equation 8

The effect of V on T:

$$ACE_{V \text{ on } T} = E[T|S = s_n, V = v_n + 1] - E[T|S = s_n, V = v_n]$$
$$= r_0 + bS + c(V + 1) - (r_0 + bS + cV)$$
$$= c$$

Equation 9

The effect of T on W:

$$ACE_{T \text{ on } W} = E[W|T = t_n + 1, X = x_n] - E[W|T = t_n, X = x_n]$$
$$= r_0 + e(T + 1) + iX - (r_0 + eT + iX)$$
$$= e$$

Equation 10

The effect of X on W:

$$ACE_{X \text{ on } W} = E[W|T = t_n, X = x_n + 1] - E[W|T = t_n, X = x_n]$$
$$= r_0 + eT + i(X + 1) - (r_0 + eT + iX)$$
$$= i$$

Equation 11

The effect of W on Y:

$$ACE_{W \text{ on } Y} = E[Y|W = w_n + 1, X = x_n] - E[Y|W = t_n, X = x_n]$$
$$= r_0 + h(W + 1) + jX - (r_0 + hW + jX)$$
$$= h$$

Equation 12

The effect of X on Y:

$$ACE_{X \text{ on } Y} = E[Y|W = w_n, X = x_n + 1] - E[Y|W = t_n, X = x_n]$$
$$= r_0 + hW + j(X + 1) - (r_0 + hW + jX)$$
$$= j$$

Equation 13

The effect of S on Z:

$$ACE_{S \text{ on } Z} = E[Z|S = s_n + 1, T = t_n, V = v_n, W = w_n, Y = y_n, X = x_n]$$
$$- E[Y|S = s_n, T = t_n, V = v_n, W = w_n, Y = y_n, X = x_n]$$
$$= r_0 + a(S + 1) + dT + fV + gW + lY + kX - (r_0 + aS + dT + fV + gW + lY + kX)$$
$$= a$$

Equation 14

The effect of T on Z:

$$ACE_{T \text{ on } Z} = E[Z|S = s_n, T = t_n + 1, V = v_n, W = w_n, Y = y_n, X = x_n]$$
$$- E[Y|S = s_n, T = t_n, V = v_n, W = w_n, Y = y_n, X = x_n]$$
$$= r_0 + aS + d(T + 1) + fV + gW + lY + kX - (r_0 + aS + dT + fV + gW + lY + kX)$$
$$= d$$

Equation 15

The effect of V on Z:

$$ACE_{V \text{ on } Z} = E[Z|S = s_n, T = t_n, V = v_n + 1, W = w_n, Y = y_n, X = x_n]$$
$$- E[Y|S = s_n, T = t_n, V = v_n, W = w_n, Y = y_n, X = x_n]$$
$$= r_0 + aS + dT + f(V + 1) + gW + lY + kX - (r_0 + aS + dT + fV + gW + lY + kX)$$
$$= f$$

Equation 16

The effect of W on Z:

$$ACE_{W \text{ on } Z} = E[Z|S = s_n, T = t_n, V = v_n, W = w_n + 1, Y = y_n, X = x_n]$$
$$- E[Y|S = s_n, T = t_n, V = v_n, W = w_n, Y = y_n, X = x_n]$$
$$= r_0 + aS + dT + fV + g(W + 1) + lY + kX - (r_0 + aS + dT + fV + gW + lY + kX)$$
$$= g$$

Equation 17

The effect of Y on Z:

$$ACE_{Y\ on\ Z} = E[Z|S = s_n, T = t_n, V = v_n, W = w_n, Y = y_n + 1, X = x_n] - E[Y|S = s_n, T = t_n, V = v_n, W = w_n, Y = y_n, X = x_n]$$
$$= r_0 + aS + dT + fV + gW + l(Y + 1) + kX - (r_0 + aS + dT + fV + gW + lY + kX)$$
$$= l$$

Equation 18

The effect of X on Z:

$$ACE_{X\ on\ Z} = E[Z|S = s_n, T = t_n, V = v_n, W = w_n, Y = y_n, X = x_n + 1] - E[Y|S = s_n, T = t_n, V = v_n, W = w_n, Y = y_n, X = x_n]$$
$$= r_0 + aS + dT + fV + gW + lY + k(X + 1) - (r_0 + aS + dT + fV + gW + lY + kX)$$
$$= k$$

Equation 19

The regression coefficients a, b, c, d, e, f, g, h, i, j, k, and l are the direct effects of the variables S, T, V, W, X, and Y on the outcomes T, W, Y, and Z as shown in Figure A1 as the pathway coefficients.

Total causal effects

The variables S, V, and X are only affected by the exogenous variables U_S, U_V, and U_X respectively.

Total effect of S on T:

$$T = bS + cV + U_T$$
$$= bS + U$$

The total effect of S on T is the same as the direct effect.

Total effect of V on T:

$$T = bS + cV + U_T$$
$$= cV + U$$

The total effect of V on T is the same as the direct effect.

Total effect of T on W:

$$W = eT + iX + U_W$$
$$= eT + U$$

The total effect of T on W is the same as the direct effect.

Total effect of X on W:

$$W = eT + iX + U_W$$
$$= iX + U$$

The total effect of X on W is the same as the direct effect.

Total effect of W on Y:

$$Y = hW + jX + U_Y$$
$$= hW + U$$

The total effect of W on Y is the same as the direct effect.

Total effect of X on Y:

$Y = hW + jX + U_Y$

$= h(eT + iX + U_W) + jX + U_Y$

$= heT + hiX + hU_W + jX + U_Y$

$= (hi + j)X + heT + hU_W + U_Y$

$= (hi + j)X + U$

$= \tau_X X + U$

The total effect of X on Y is τ_X Equation 20

Total effect of S on Z:

$Z = aS + dT + fV + gW + lY + kX + U_Z$

$= aS + d(bS + cV + U_T) + fV + g(eT + iX + U_W) + l(hW + jX + U_Y) + kX + U_Z$

$= aS + dbS + dcV + dU_T + fV + ge(bS + cV + U_T) + giX + gU_W + lh(eT + iX + U_W) + ljX + lU_Y + kX + U_Z$

$= aS + dbS + dcV + dU_T + fV + gebS + gecV + geU_T + giX + gU_W + lheT + lhiX + lhU_W + ljX + lU_Y + kX + U_Z$

$= aS + dbS + gebS + lhe(bS + cV + U_T) + lhiX + lhU_W + ljX + lU_Y + kX + U_Z + dcV + dU_T + fV + gecV + geU_T + giX + gU_W$

$= aS + dbS + gebS + lhebS + lhecV + lheU_T + lhiX + lhU_W + ljX + lU_Y + kX + U_Z + dcV + dU_T + fV + gecV + geU_T + giX + gU_W$

$= (a + db + geb + lheb)S + U$

$= \tau_S S + U$

The total effect of S on Z is τ_S Equation 21

Total effect of T on Z:

$Z = aS + dT + fV + gW + lY + kX + U_Z$

$= aS + dT + fV + g(eT + iX + U_W) + l(hW + jX + U_Y) + kX + U_Z$

$= aS + dT + fV + geT + giX + gU_W + lh(eT + iX + U_W) + ljX + lU_Y + kX + U_Z$

$= aS + dT + fV + geT + giX + gU_W + lheT + lhiX + lhU_W + ljX + lU_Y + kX + U_Z$

$= (d + ge + lhe)T + aS + fV + giX + gU_W + lhiX + lhU_W + ljX + lU_Y + kX + U_Z$

$= \tau_T T + U$

The total effect of T on Z is τ_T Equation 22

Total effect of V on Z:

$Z = aS + dT + fV + gW + lY + kX + U_Z$

$= aS + d(bS + cV + U_T) + fV + g(eT + iX + U_W) + l(hW + jX + U_Y) + kX + U_Z$

$= aS + dbS + dcV + dU_T + fV + ge(bS + cV + U_T) + giX + gU_W + lh(eT + iX + U_W) + ljX + lU_Y + kX + U_Z$

$= aS + dbS + dcV + dU_T + fV + gebS + gecV + geU_T + giX + gU_W + lhe(bS + cV + U_T) + lhiX + lhU_W + ljX + lU_Y + kX + U_Z$

$= aS + dbS + dcV + dU_T + fV + gebS + gecV + geU_T + giX + gU_W + lhebS + lhecV + lheU_T + lhiX + lhU_W + ljX + lU_Y + kX + U_Z$

$= (dc + f + gec + lhec)V + aS + dbS + dU_T + gebS + geU_T + giX + gU_W + lhebS + lheU_T + lhiX + lhU_W + ljX + lU_Y + kX + U_Z$

$= \tau_V V + U$

The total effect of V on Z is τ_V Equation 23

Indirect causal effects

Indirect effect of X on Y = total effect - direct effect
$= \tau_X$ - j
= hi + j - j
= hi Equation 24

Indirect effect of S on Z = total effect - direct effect
$= \tau_S$ - a
= a + db + geb + lheb - a
= db + geb + lheb Equation 25

Indirect effect of T on Z = total effect - direct effect
$= \tau_T$ - d
= d + ge + lhe - d
= ge + lhe Equation 26

Indirect effect of V on Z = total effect - direct effect
$= \tau_V$ - f
= dc + f + gec + lhec - f
= dc + gec + lhec Equation 27

References

[1] Pearl, J., & Mackenzie, D. (2018). The book of why: the new science of cause and effect. Basic books.
[2] Glymour, M., Pearl, J., & Jewell, N. P. (2016). Causal inference in statistics: A primer. John Wiley & Sons.
[3] Pearl, J. (2009). Causality. Cambridge university press.
[4] Van Rossum, G., & Drake Jr, F. L. (1995). Python tutorial (Vol. 620). Amsterdam, The Netherlands: Centrum voor Wiskunde en Informatica.
[5] Sharma, A., & Kiciman, E. (2020). DoWhy: An end-to-end library for causal inference. arXiv preprint arXiv:2011.04216.